Genomics 98 (2011) 242–252

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Perspective

Decoding the genome beyond sequencing: The new phase of genomic research

Henry H.Q. Heng [a,b,c,*], Guo Liu [a], Joshua B. Stevens [a], Steven W. Bremer [a], Karen J. Ye [a], Batoul Y. Abdallah [a], Steven D. Horne [a], Christine J. Ye [d]

[a] Center for Molecular Medicine and Genetics, Wayne State University School of Medicine, Detroit, Michigan 48201, USA
[b] Karmanos Cancer Institute, Wayne State University School of Medicine, Detroit, Michigan 48201, USA
[c] Department of Pathology, Wayne State University School of Medicine, Detroit, Michigan 48201, USA
[d] Department of Internal Medicine, Wayne State University School of Medicine, Detroit, Michigan 48201, USA

ARTICLE INFO

Article history:
Received 4 March 2011
Accepted 18 May 2011
Available online 26 May 2011

Keywords:
Genome theory
Genome context
Genome chaos
Genome instability
Non clonal chromosome aberrations (NCCAs)
Clonal chromosome aberrations (CCAs)
Genetic coding
Punctuated cancer evolution

ABSTRACT

While our understanding of gene-based biology has greatly improved, it is clear that the function of the genome and most diseases cannot be fully explained by genes and other regulatory elements. Genes and the genome represent distinct levels of genetic organization with their own coding systems; Genes code parts like protein and RNA, but the genome codes the structure of genetic networks, which are defined by the whole set of genes, chromosomes and their topological interactions within a cell. Accordingly, the genetic code of DNA offers limited understanding of genome functions. In this perspective, we introduce the genome theory which calls for the departure of gene-centric genomic research. To make this transition for the next phase of genomic research, it is essential to acknowledge the importance of new genome-based biological concepts and to establish new technology platforms to decode the genome beyond sequencing.

1. Introduction

Since the inception of the journal *Genomics* nearly 25 years ago, rapid development of various genomic technologies has greatly advanced the science of genomics. However, despite cutting edge technologies including whole genome scanning [1], global gene expression profiling [2], copy number variation analysis [3] and massive parallel sequencing [4], the understanding of the human genome and the mechanism of human diseases remains a challenging process [5–7]. These powerful technologies have generated scores of data, which paradoxically challenge the framework of current genomics and gene based concepts of common disease, including the rationale of analyzing large numbers of diverse samples with the highest resolution possible. Many diseases are in fact system diseases where sundry genetic variations can be involved in a seemingly stochastic fashion. Furthermore, heterogeneity occurring at multiple levels is a key feature of these diseases (rather than just "genetic noise"), which cannot be addressed simply by sequencing DNA and increasing the sample size [8]. This issue represents the very reason for the failure to identify major causative genes/variants in many common diseases including cancer despite extensive large scale sequencing and whole genome scanning [5].

There are two obvious but somewhat contrary options that can be undertaken to move the field forward. One popular option is to continuously push the limits of technology by increasing the resolution and speeds while lowering costs in order to analyze more samples [4,9,10]. It is believed that studying a larger number of samples will yield the long anticipated genetic patterns of disease by the elimination of 'noise'. Unfortunately, many initial reports of this approach have generated contradictory conclusions, revealing enormous diversity rather than the expected reduction in diversity, and that high levels of genetic heterogeneity seem to be the general rule [5,11,12]. Questions are now being raised about whether data from large scale genomic studies will ever prove to be of promised clinical value, even if each personal genome or "cancer genome" is sequenced [13,14]. The extreme complexity of disease heterogeneity, encompasses the following: low penetrance of specific gene mutations within patient populations; multiple genetic–epigenetic and environmental interactions; and the influences of stochastic evolutionary processes, render most individual molecular mechanisms less than useful for clinical prediction.

A second, new option requires a drastic change in our thinking, and is a departure from the traditional genetic framework and will provide answers from a different perspective or level of genetic organization rather than mainly focusing on DNA and RNA sequences. The basis for this option is that genome alterations are more common and profound than individual gene mutations in most human disease conditions. This new conceptual framework based on the genome

* Corresponding author at: Center for Molecular Medicine & Genetics, Wayne State University School of Medicine, 540 E. Canfield, 3226 Scott Hall, Detroit, MI 48201. Fax: +1 313 577 5218.
E-mail address: hheng@med.wayne.edu (H.H.Q. Heng).

0888-7543/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2011.05.008

theory calls for a redirect of our efforts to systematically decode genetic information stored at the genome level [7,15,16]. Initial calls for just such a change have emanated from the field of cancer research [5,7,17–26].

We strongly support the option of refocusing on the entire genome (not just the sequence of), not only because this approach has been overlooked, but also because of its ultimate importance in understanding how the entire genome functions and in the underpinnings of the general mechanism of many common complex diseases [7]. In this perspective, we will briefly discuss the key differences between genes and the genome by walking readers through our own experience of making the transition from the gene centric view to the genome theory [27]. In particular, to convince readers of the importance of this issue, we would like to point out that current genome study efforts only decode parts of the genome and do not address the key issue of decoding the genome as a whole system. Even for the ENCODE project (the encyclopedia of DNA elements) and Human Epigenome project (as well as many other 'omics' projects) [28,29], the conceptual framework is still the gene theory. At a fundamental level, DNA sequences (including their chemical modifications) and the genome represent distinctively different levels of coding and system control, and future genomic research must directly address the issue of genome coding, genome system control, and how interactions work across different genetic and epigenetic levels [8]. Equally important, new technical platforms are urgently needed to synthesize information at the higher levels and integrate them with the genome system. Emergent properties of the genome suggest that system information at the genome level is not a simple summary of gene sequence information. New technologies also need to integrate other key features of normal and diseased genomes such as: heterogeneity at multiple levels; differences between the system status (such physiological and pathological conditions where pathological conditions often involve genome alterations) and stochasticity of somatic evolution.

2. Gene vs. genome

2.1. The genome is not equal to the sum of all genes or its entire sequence

The genome is the entity containing an organism's hereditary information and the main evolutionary selection platform [27,30]. Traditionally, heritable information has been thought to be encoded exclusively in DNA and RNA sequences. The current, popular concept of the genome where the collection of all genes and non-coding sequences explain a given species has been influenced by the gene centric concept. A key unique feature of the genome however, is the genomic topology (a multi-dimensional interactive relationship that exists between genes and is the physical basis of genome architecture) and the emergent properties that exist at this higher level which have been largely ignored. As a result, the terminology, the Human Genome Project, as used by sequencing consortiums, implies that decoding DNA is equal to decoding the genome. This has spawned many popular but incorrect analogies, including considering the genome to be a book, where each chromosome represents a chapter. The problem with this metaphor is that one cannot simply read the basics of each chapter and comprehend it without including the multi-dimensional interactions within the system. A chromosome does not stand on its own as a biological entity and therefore there are no meaningful messages based on individual chromosomes. To put it succinctly, a simple parts list does not give a clue as to the assembly instructions. Similarly, the conventional statement that the sequencing of the human genome has provided a roadmap (or the foundation) of modern biomedical research is flawed. It is flawed particularly with regard to nonlinear systems as most complex systems have multiple levels of organization and the nonlinear relationship between them is connected through emergent properties, which is difficult to understand by only summarizing information of the lower parts. The reductionist tradition of understanding the "parts" first before understanding the "whole" system is only effective in a linear system [15,31].

Clearly, considering the relationship of the parts (genes) versus the whole (genome), where the whole is more than the sum of all the parts, and includes the 3-D interactive structure of the genome within nuclei, the current reductionist approach of treating the genome as a "bag of genes" or a collection of linear DNA structures does not mirror the complexity of the genome system. To illustrate this point, we have introduced the term, 'genome context', to differentiate the gene and the genome [21,27]. Genome context (the DNA sequence plus the genomic topology), rather than gene content, defines the structure of a genetic network and is the total interactive package that functions in organismal and somatic cell evolution. The revelation of the genome context results in the need for additional, crucial questions to be asked within the genomic community. These include: When a genome is altered, does the same gene mutation have the same biological meaning as it does in the original, unaltered genome? If the genome's key properties are emergent from the DNA level and are fundamentally different from DNA itself, is DNA sequencing crucial to understand the function of a eukaryotic genome?

In the real world the difference between a parts list and the blueprint of parts assembly is clear. In genome research, however, this key difference is often forgotten. It is thus necessary to separate genes and the genome as these are two different entities in genomic research. The difference between them is not just a difference of quantity but also a difference of the level of organization. One interesting analogy is to consider genes as building material, parts or even tools, and the genome as the architecture. The same materials can be used to construct different architectures with distinct functions [21]. As pointed out by Barbara McClintock, the genome is the organization that is responsible for activating and reconstructing in response to environmental challenges [32].

It is true that increasing attention has been given to the "gene context" rather than genes alone. However, without the conceptual framework of the genome theory, isolated approaches will not solve most paradoxes that the gene theory created [27].

2.2. Different genetic coding and control systems

If the above analogy that the gene/genome relationship represents a part versus the whole relationship is correct, then it is obvious that genes and genomes represent different levels of the system with different mechanisms for coding genetic information. Understanding how DNA codes RNA and proteins has been a major achievement of molecular biology [33], however, the codes for producing the parts using DNA and the codes for assembling parts using the package of the genome are very different. Current systems biology suggests that the function of the genome depends on the genetic network, without knowing what defines the genetic network structure in the first place. In contrast, genome theory states that the genome context defines the structure of a genetic network by creating the structural or topological basis of genetic interactions [21,27], thus a specific network is the emergent property of a given genome (Fig. 1).

According to the genome theory, both genes and the genomic topology are the key to maintaining the genome system. Therefore, new systems can be effectively created naturally either by creating many novel genes, or by re-organizing the existing genome through changing the genomic topology rather than gene content. Many less complex species like sponge and water fleas have comparable or even more genes than humans (comprised of a simple body plan without organs, muscles or nerve cells, the sea sponge has 18,000 genes compared to 21,000 genes for humans, while Daphinia water fleas have over 30,000 genes) [34,35], and most mammals share similar genes but have very different chromosomal arrangement and

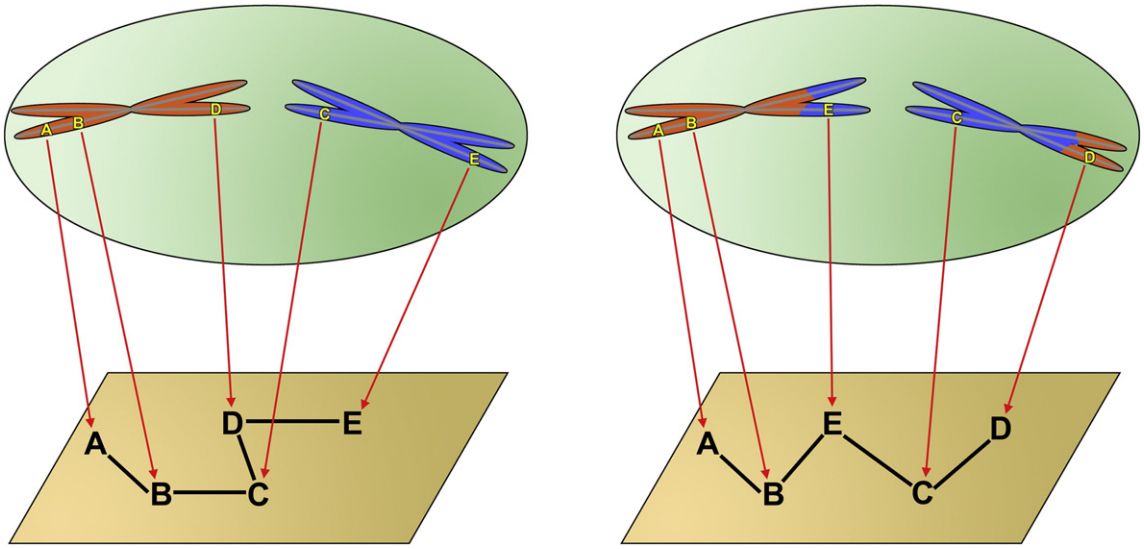

Fig. 1. A diagram demonstrating how changes of genome topology alter the structure of protein networks (diagram modified from Heng [27]). For simplicity, only two chromosomes are drawn within the nucleus representing the genome. Genes are designated A, B, C, D, and E within the chromosomes. When a chromosomal translocation occurs, the genome topology is altered affecting the physical relationship between chromatin domains, which changes the overall genetic network structure. As a result, the protein network changes illustrated by the changed relationship between proteins A, B, C, D, and E. DNA codes the individual genes or parts and the interactive relationship among all genes and non-coding sequences are achieved by the 3-D genome topology which serves as the platform or matrix where the self-organization principle works to form the genetic network.

composition (genome topology) [36–38]. We therefore conclude that reorganizing the genome is more influential than inventing new genes in terms of creating new systems.

It has been puzzling how the system preserves or codes genetic information at levels above DNA such as the structure of a genetic network. A recent effort to re-interpret the main function of sex has unexpectedly provided the answer to this dilemma. It turns out that both gene content and genome topology are ensured by sexual reproduction, where the identical genome (reflected as karyotype) is a key and any significantly altered genome will be eliminated by the sexual filter [39,40]. As the same karyotype ensures the same genome topology, and sexual reproduction maintains the same genome (karyotype), sex in fact preserves the structure of a genetic network by maintaining the genome context. Now we understand why in all eukaryotes genetic information is transmitted from cell to cell by chromosomes and that the genome context defines a species. In asexual species lacking the sexual filter genome dynamics are extremely high, resulting in constant alteration of the genome system. Similarly, very rapid karyotypic evolution can occur during somatic cell evolution when there is no sexual filter to preserve the genome, which might be the reason why cancer evolution can be achieved in a relatively short time window within the human body while organismal evolution with species stasis requires millions of years [24]. The important message here is that the genetic information of system dynamics is not entirely encoded within the DNA. The coding information at the gene level is conserved across species, but the coding for gene interaction at the genome level is highly dynamic among species while it is highly conserved within the same species.

As for the issue of whether genes/genomes represent different control systems, the answer is obvious. No gene is an island and most genes are not independent information units. Despite each gene encoding a specific protein or RNA, global interactions are ultimately controlled by the higher genome level, and the genome is a package where emergent properties rule by balancing out contributions of individual genes. This balance occurs as there are so many genes contributing to the emergent properties, the influence of individual genes is limited by the genome context. Interestingly, there are many examples that illustrate the conflicting relationship between DNA and the genome level. When identical transgenic DNA molecules are introduced into mouse meiotic chromosomes, the loop size is influenced by the integration sites along the chromosome. These inserts form shorter loops integrating close to the telomeric region rather than being inserted at the middle of chromosomes, demonstrating chromosomal position constraints of DNA's behavior [41]. Similar observations include studies on position effects [42]. Similarly, when multiple copies of genes are integrated into the mouse genome in a tandem array, only one copy is expressed possibly due to the chromatin loop constraint, demonstrating that an organizational role exists at the level above the gene [43–45].

Perhaps the most striking example of the conflict between the gene and genome level is the dual functions of meiosis to reduce and also promote genetic variation. Traditionally, sexual reproduction has been thought to be a process that increases genetic diversity through the mixing of genes, an obvious advantage of meiosis. Application of genome theory, however, has surprisingly demonstrated that the main function of sexual reproduction is to reduce genetic diversity at the genome level and maintain system identity, while imparting a certain degree of diversity at the gene level [39]. This conclusion is supported particularly through the re-examination of the evolutionary function of meiosis [40]. According to Wilkins and Holliday's recent insightful analysis, "…the conclusion is surprising: the initial function of chromosome pairing was to limit, not enhance, recombination". To further reconcile the field, the main function of sexual reproduction has been linked to the reduction of genetic diversity in general and at the genome level in particular [46]. Interestingly, the gene/genome relationship also might resolve the conflict between short term evolutionary adaptation and long term system stasis [47], as the genome functions as a main evolutionary constraint (46, Heng, unpublished data). The concept that different levels of a genetic system need different coding systems could have broader implications. It strongly suggests that when considering a multiple level complex system, one should not rely on coding or information at the lower level to explain the behavior of the higher level, as distinct coding and control systems are involved.

2.3. The Genome Theory and its clinical importance

To advance the field, the genome theory of organismal and somatic cell evolution has been proposed based on key differences between genes and the genome [24,27] (Fig. 1 and Table 1). The key message of

Table 1
Key differences between genes and the genome.

Gene	Genome
Coded by DNA/RNA	Coded by entire set of chromosomes
Information preserved by the same DNA sequences	Information preserved by the genome context
Decoding by sequencing DNA	Decoding by illustrating 3-D interaction of parts
Information on the parts/tools (parts list)	Information for system assembly/ function (blueprint)
Status change may lead to network rewiring	Alteration creates a new genome system
Involves micro-evolution (changes are often neutral)	Usually involves macro-evolution (changes are often harmful)
Modifies a system with new features	Defines a system with new gene context and potential
Forms specific RNA/protein Conserved across species	Defines a genetic network Conservation of a species

the genome theory is that future genomic research should refocus on/at the genome level, as the genome package defines the genetic network and its potential response towards environmental stresses [27,32]. The goal of sequencing the genome has been achieved and it is now time to move to the next level of understanding of the genome system rather than to continue to compile data on its parts list. A switch to genome based approaches is essential to apply basic genomic research to the clinic, as genome level changes often drive DNA level changes.

Recent studies demonstrate that a few chromosomal translocations often couple with hundreds of rearrangements at the DNA level [48]. We have compared the copy number variations and karyotypic dynamics (in the form of frequencies of NCCAs) during spontaneous cellular transformation in mouse cells, and the highest level of copy number variation occurs in association with the highest level of non-clonal chromosome aberrations (NCCAs). This suggests that when the system is unstable, elevated numbers of genomic alterations can be detected across different levels of the system, though change at the karyotype level typically reigns supreme (unpublished data). In fact, the de novo locus-specific rate of genomic rearrangement is at least 100- to 10,000-fold greater than the rate of point mutations [49] and there are an increasing number of reports linking many common diseases to genome alterations [7,17,21,50]. Altogether, this mounting evidence calls for a re-examination of the link between stochastic genome alterations and common diseases especially cancer [7,24]. Linking more diseases to genome alterations rather than specific gene mutations will ultimately challenge the validity of the current sequencing efforts being used in an attempt to understand the general mechanism of common diseases such as cancer. The influence of different gene level alterations that can lead to the same disease reduces the importance of monitoring any particular individual gene. Thus, new genomic based strategies are clearly needed to predict likely disease potential.

3. Gene and genome based technologies

3.1. Most current genomic technologies are based on the concept of average profiles

As illustrated by this special issue, the vast majority of molecular profiling methods are based on studying an average profile of cells. In addition to the rationale of using the average to wash out 'noise', enabling biological pattern recognition, technical limitations are another reason why mixed cells are often used in molecular analysis. If cell populations were truly homogeneous this method would be justified however, for most disease conditions, compromise of homeostasis of the system, leads to high heterogeneity. As heterogeneity is not insignificant noise, but a key feature of many diseases the use of profile averaging is misleading [8,31]. For example, an average profile tells little about potential drug resistance, because heterogeneity predominates in these situations making the use of averaging technologies erroneous [16]. Consider two populations of cells descendent from healthy cells with a normal chromosome content. Each of the populations undergoes chromosome gain, and thus differs from the normal parental cells. In the first homogeneous population, every cell contains the exact same chromosomal change, so that the population average is equal to any individual cell. In the second highly heterogeneous population changes occur that result in a population average that is the same as the first population, however each individual cell has a unique number of chromosomes, and thus different genome system. A drug could be designed to effectively eliminate the average population, and this drug would work well in the case of the first population, but it may have little or no effect on the second heterogeneous population. Similarly, when a given drug is used, the average effect of the drug can be monitored using a Western blot analysis that detects the proteins involved in cell death pathways and measures average pathway response. However, when individual cells are analyzed, there often are some exceptions noted including cells that display even better growth. This type of heterogeneity certainly contributes to drug resistance.

The idea of averaging profiles has been problematic for genetic validation as well. If there are no average mutations, how can we validate different mutations? Similarly, if the majority of cells are different in terms of a gene mutation profile, what is the profile average we are measuring? This point has been forcefully illustrated time and time again in cancer research where a high degree of heterogeneity is a common feature at the genome level [18–20]. Similarly, at the gene level, methods of direct mutation detection rather than profile averaging have shown that there are large numbers of random mutations within each tumor, the majority of which are not recurrent (thus not detected by averaging methods) [51]. Indications are that the development and application of single cell profiling methods used in genomic studies are increasing, which represents a positive future direction.

3.2. How can stochastic genome alterations be monitored?

To understand the dynamic regulation of a genetic network, at least three types of players need to be analyzed including genes, the genomic topology and how the self-organization principle actually works within the nucleus. Currently, most efforts have been focused on the mechanism of network rewiring, without considering the genomic topology and how system emergence occurs from lower level components [52]. Therefore, we need to be aware that there might be a huge difference between the current knowledge of genetic networks and the genomic reality. Even with simplified versions of current analysis, there are two main aspects of network dynamics: one is rewiring through small scale modifications such as gene mutation, epigenetic regulation and copy number variation. This is the main focus of current systems biology applied to analyzing network regulation [53,54]. Another is system recreation through reorganization of the genome using the same genes, a key topic that is still ignored [8,21,25]. In organismal evolution, both processes occur with genome reorganization being the more dominant for speciation. Based on the genome theory, most diseases will involve the latter category as pathological conditions are often caused by large scale genome alterations.

Clearly, we do not have all the technologies available yet to address these issues in a systematic way. There however have been some important developments, which have started to move in this direction. The following are some interesting examples.

3.2.1. Mapping and sequencing interaction sites

The importance of chromosomal and genome topology has long been appreciated by molecular cytogenomics [17,21,55] and includes the use of chromatin loop domains to study the distribution pattern of specific regions of chromosomes within a nucleus [56], and pinpoint neighboring chromosomes that are likely to involve translocations [57]. It is now clear that chromosomes are arranged differently in different cell types, and that the 3-D chromosomal position influences the gene activities [58]. However, such important messages have failed to capture the interest of main stream genome research, possibly due to limited resolution and inability to apply this type of research in a high throughput manner [21]. Recently, using proximity-based ligation (cross-linking to capture the information of 3-D interaction among chromatin domains) and massive parallel sequencing, the interaction sites of the genome were studied. Long-range interactions revealed folding principles of the human genome. Interestingly, this study confirms many theories generated from traditional cytogenetic studies, such as the presence of chromosome territories and the spatial proximity of small, gene-rich chromosomes within the nucleus, and further illustrates an additional level of genome organization that is characterized by the spatial segregation of open and closed chromatin to form two genome-wide compartments [59]. Various high-C analyses have been performed [60], for example, a 3-D model of a 500-kb region of human chromosome 16 has been built [61]. We hope that such an approach will soon capture more interest in the field [62].

In regards to evolutionary relationships, it is now a common practice to compare sequence conservation of a specific gene or sequence across different species. Future genomic comparisons are needed to study the conservation of interactive relationships among networks by comparing the genome context. Among the potential limitations of imparting gene theory to explain genome issues which involve a different level of the system. For example, when studying specific chromosomal translocations in cancer, focus has centered on identifying genes altered by translocation (dysfunctional genes or newly formed fusion genes), without realizing that in addition to these directly changed genes, the more dominant effect might come from the alteration of the entire genomic topology defined network structure. The identification of a fusion gene might only represent part of the story even though the fusion gene is what we can easily trace based on current technologies. Considering that there is a high degree of heterogeneity between detectable fusion genes and cancer phenotypes in most cancers and progression is diverse in patients with the same translocation, more emphasis should be placed on the overall genome alteration. Accordingly, utilization of the genome based framework needs to be sped up to help guide and benefit future technological developments.

3.2.2. Using non clonal chromosome aberrations to study genome instability

Karyotype analysis has been extensively utilized in cancer and human genetics research. Cytogenetic markers not only have helped to identify disease specific genes, but have also been used in clinical diagnosis [55,63]. Traditionally, however, only clonal chromosome aberrations are reported, particularly when commonly shared among patients. In contrast, detected NCCAs have been considered to be insignificant genetic noise, and thus ignored despite occurring commonly [17–20]. According to the genome theory, the level of seemingly random genome alterations is not noise, but represents system heterogeneity. Using cancer evolution as a model, we have linked increased frequencies of NCCAs to system dynamics and population diversity. Interestingly, the level of NCCAs can directly reflect system instability, regardless of the causes of the NCCAs.

There is a seemingly unlimited list of factors that can lead to system instability under certain conditions. In addition to genes that are known to maintain genome integrity, many onco-proteins, carcinogenic treatments, viral infections, inflammation, the aging process, various disease conditions and even general physical stresses can contribute to system instability that is linked with elevated frequencies of NCCAs [18]. Furthermore, following the comparison of different cancer models with different molecular mechanisms, we found that the general mechanism can be explained by the direct relationship between stress induced system dynamics (illustrated by elevated NCCAs) and tumorigenicity [64]. By connecting the dots, we have finally established the evolutionary mechanism of cancer.

$$\text{Evolutionary mechanism} \geq \sum \text{all individual molecular mechanisms}$$

The evolutionary mechanism of cancer can be explained by three simple key components, namely, stress induced system dynamics, population diversity and genome (not gene) mediated system macro-evolution. To state that the evolutionary mechanism of cancer is equal to the collection of all individual molecular mechanisms, puts individual mechanisms that most researchers are working on into perspective, as each specific mechanism can explain some but not all cases [24,64]. Since there are countless possible stochastic interactions that could occur between known molecular mechanisms, and as evolution is an emergent property, the evolutionary mechanism is actually much larger than the collection of all molecular mechanisms and this emphasizes that new strategies to monitor evolutionary potential are now needed rather than strategies that focus only on individual molecular mechanisms. In fact, the initial efforts of The Cancer Genome Atlas have revealed that most key cancer gene mutations have low penetration among patients. An even greater challenge is not just characterizing each genetic aberration's defined pathway that can contribute to cancer (this can be achieved by sequencing a great number of samples), but more important is predicting the pathway switching that occurs during treatment. This can further reduce the clinical significance of individual genetic aberrations when the penetration is very low within a patient population.

Interestingly, NCCAs are elevated in diseases other than cancer. For example, we have detected much higher frequencies of NCCAs in a limited number of Gulf War Syndrome patients (unpublished data). Similarly, various mouse models of metabolic diseases have been shown to exhibit increased NCCAs. It is thus necessary to evaluate more patient samples to investigate the connection between system instability and disease [7,31].

3.2.3. Analyzing genome chaos

To illustrate how drastically the genome can be re-organized under stress, forming a totally new system, discussion of the phenomenon of genome chaos (or karyotypic chaos) is required. Our in vitro immortalization model has illustrated that the genome displays massive genome alterations (including numerical and structural alterations) and this process is essential for cellular immortalization. Many translocation events occur in each genome, and as every cell is different, this phenomenon provides a large variety of combinations of genetic material for genome evolution. Genome chaos can be observed from an array of experimental and pathological conditions [18]. Fig. 2 illustrates one example of genome chaos induced by drug treatment. SKY technology has been used to illustrate the complexity of the translocations, as each color represents one chromosomal origin. Many translocated chromosomes can be rapidly formed from fragments of other chromosomes. It is worth noting that SKY or multiple color FISH methods are essential to study these massively altered karyotypes. Initially these technologies were developed as a powerful method to identify the pattern of karyotype abnormalities in cancer [64–66], and now our research

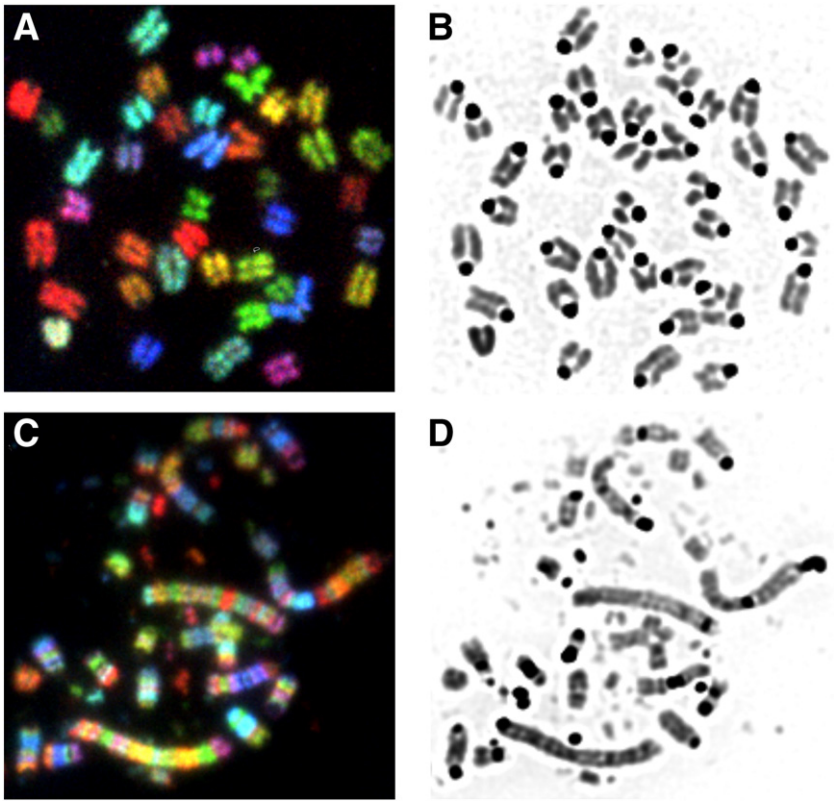

Fig. 2. SKY (spectral karyotype) image of genome chaos where there are massive translocations among chromosomes. A. In a normal mouse genome, each chromosome is "painted" by one color. B. The reverse DAPI image of the same mitotic figure in A. C. There are massive translocation events detected within a chaotic genome following short drug treatment. These newly formed giant chromosomes are possibly derived from complex chromosomal fusion, with each color representing their chromosomal origin. D. The reverse DAPI image of the same mitotic figure in C. Drastically altered genomes represent a new genome system and understanding the mechanism generating genome chaos and genome evolution can provide insight into genome re-organization.

has demonstrated its ultimate usefulness in the study of stochastic genome alterations such as NCCAs.

That so many genome level alterations can be generated so rapidly when a cell endures stress is very important to understand the mechanism of genome reshuffling and its evolutionary implications, particularly in understanding the mechanism of drug resistance. It is possible that chromosome fragmentation, which is a key mechanism of mitotic cell death, plays a role in generating genome chaos. Chromosome fragmentation is different from apoptosis and represents a powerful means to cut chromosomes into different pieces (Fig. 3). Our hypothesis is that, following incomplete chromosomal fragmentation [17,21,67,68], there is a new mechanism to rejoin the massive amount of fragments to form chaotic chromosomes within a few generations. Due to this massive re-organization occurring in such a short time, some unknown mechanism rather than conventional DNA repair must be involved. Interestingly, similar phenomena have been reported in clinical samples where massive genome alterations were detected which involved a limited number of chromosomes. The term "chromothripsis" has been used to describe this specific form of genome chaos [69,70]. Systematic analysis is needed to correlate which massively altered genomes (different systems) are associated with which different network structures. Sequencing the newly formed rejoined chromosomes is also of interest.

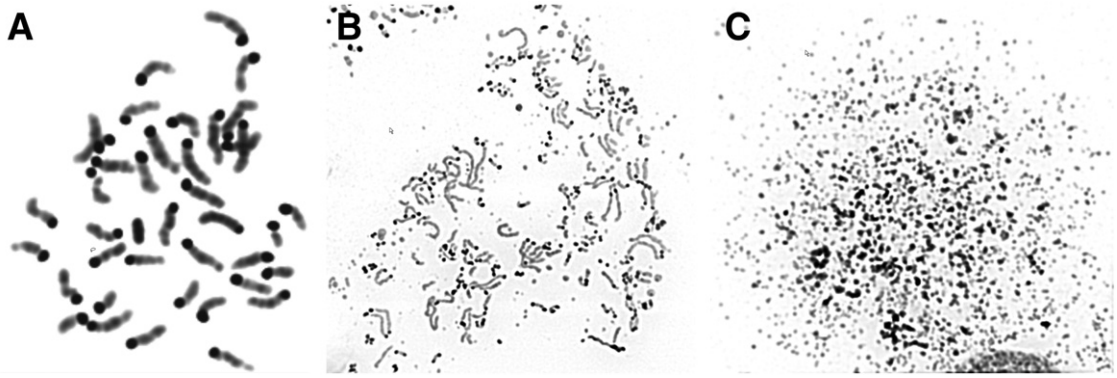

Fig. 3. Images of chromosome fragmentation where mitotic chromosomes are cut into small pieces. A. Reverse DAPI image of a normal mouse karyotype. B. Example of chromosome fragmentation where both chromosomes and fragments are detected. C. An example of late stage chromosome fragmentation where there is minimal visible chromosomal morphology. Chromosome fragmentation is a newly identified form of mitotic cell death. It is possible that chromosome fragmentation contributes to the formation of genome chaos. Chromosome fragmentation also represents another example of previously disregarded non clonal chromosome aberrations.

4. Evolutionary and system approaches

New technologies need to be built around both holistic system approaches and evolutionary analysis, as genomes of different species are the result of genomic evolution, and most diseases are a system problem and disease initiation, progression and drug response/resistance all represent typical evolutionary processes. Without the evolutionary context, much of our genomic knowledge does not make sense. For example, when discussing the issue of how much of the genome is functional and why there are so many non-coding sequences or non-essential genes, the historical evolutionary process of genome reorganization and the dynamic of environmental variation must be considered. The genomic evolution process is mainly based on genome reshuffling rather than "useful gene accumulation" and genomes are products that reflect their historical formation and their mechanism of evolutionary selection. This point has often been forgotten when searching for essential genes. The yeast model illustrates this, as about 70% of yeast genes can be deleted under experimental conditions without any apparent repercussion. This approach is not representative of real world situations and thus is fundamentally limited. In order to properly consider future challenges, the following issues are worth discussing.

4.1. Which level should be the priority of a study?

There are many genetic and epigenetic levels that can be linked to disease. At the gene level, gene mutations, splicing variations, and epigenetic regulation can be linked to gene functions. At the chromatin level, epigenetic regulation, domain configuration and 3-D interactions are important. At the genome level, there is massive copy number variation, and a great number of cytogenetic aberrations that contribute to a given phenotype. How can an experimental system be established to combine all of this information together? How can these levels be prioritized when their information is conflicting? Which levels of information are more important in terms of clinical implications? Should we pay more attention to global approaches or to reductionist approaches? Is it possible to ignore some high resolution data at the lower system levels and focus on higher system levels or even the overall system behavior?

To address these issues, one must consider both information and evolutionary theories. According to the information theory, regarding multiple levels of a system, lower level information often has less to do with system control than higher levels. Consideration of the recently established relationship between genes and the genome, and between micro and macro evolution [27], enables classification of diseases into different categories for future studies. If the issue is driven by macro-evolution such as cancer where genome system replacement is the key, the research focus must be on genome alterations. In contrast, if the research involves gene mutation without genome level changes, such as developmental processes that are not a result of changing the genome, then gene level analysis is appropriate. Therefore, the first question to consider with regard to specific diseases is determining whether macro or micro-evolution is involved.

To judge whether macro-evolution is involved, the process dynamics need to be analyzed (of a cell culture process or animal model) by comparing karyotypes. This can be achieved by spectral karyotype (SKY) analysis. Different observation time points during the process are selected and 50–100 mitotic figures are analyzed to record the patterns of NCCAs and CCAs (clonal chromosome aberrations). If there is a high level of genome turnover during the monitoring process, then macro-cellular evolution is likely involved. It should be pointed out that, there are many types of chromosomal aberrations, and many such as NCCAs have been traditionally ignored. For example, defective mitotic figures and chromosome fragmentation were regarded as slide preparation artifacts until our recent studies [17,21,71,72]. Despite our series of studies demonstrating the importance of NCCAs [18–22], the research community has been slow to accept this important development. Systematic studies are needed to characterize these ignored chromosomal aberrations. By monitoring NCCAs, it will be much easier to link many diseases with system instability and the macro-evolutionary process.

4.2. Is there a disease genome?

Diseases can be divided into four groups: based on their genetic penetration and whether or not genome instability is involved. In types A and B, genetic factors are commonly shared in patient populations with stable (A) or unstable genomes (B). For types C and D genetic factors are rare in patient populations with stable (C) or unstable (D) genomes. Many common diseases belong to the type D group [7]. If the genome is stable and there are overwhelming numbers of normal karyotypes across various stages of the disease condition, such a disease should be considered a gene based disease. If stable but abnormal karyotypes dominate during different stages of the disease progression (like Down's syndrome), then this genome would be considered to be a "disease genome". If however, the vast majority of patients display drastically different karyotypes (such as cancer patients), there is no cancer genome. Unfortunately, to date, most of the cancer genome sequencing projects have ignored the differences between genomes by focusing on known genes.

When genome-defined systems are completely different, what is the meaning of comparing the same gene mutation? Recent reports of various cancer genome sequencing projects have revealed high levels of genome alterations with hundreds of rearrangements in individual tumors. However, the majority of analyses still focus on the association of known cancer genes without illustrating key genome level contributions. Clearly, these massive genome alterations change the meaning of individual cancer gene mutations as they have created totally new genome systems [5,21,27].

The concept that specific functions of individual genes can be altered by newly formed genomes has been elegantly confirmed by yeast evolutionary studies [73]. The MYO1 gene encodes proteins essential for cytokinesis. Following MYO1 deletion, cytokinesis function was restored in some cells through extensive genome alteration. When MYO1 was re-introduced back to these altered genomes, the gene was no longer relevant to these new systems and no longer participated in the original functions. Similar thinking and approaches need to be applied to other common diseases as well, despite the difference of the degree of genome alteration between cancer and other common diseases.

4.3. Two phases of evolutionary dynamics

Cancer evolution can be dissected into a repeating series of two phases of evolutionary events [18]. In the punctuated phase, macro-evolution dominates while in the stepwise phase, there is a mixture of macro-and microevolution occurring. Knowing the interaction of the two phases of somatic cell evolution is useful to monitor system evolution and adopt suitable monitoring technologies accordingly. In stable systems, evolution can be triggered by either micro- or macro-evolutionary events, so both gene level changes and genome level changes need to be analyzed. In the dominant macro-evolution phase, gene studies are of very limited value. Significantly, cancer evolution can be observed as multiple cycles of NCCAs/CCAs which transition during key status changes such as immortalization, transformation, and acquisition of drug resistance. Therefore, the evolutionary journey throughout each cancer event (from initiation through later clinically significant stages), is not achieved by one genome system that remains through all these stages. In contrast, it is achieved by a series of genome replacements that occur throughout the macro-evolutionary process (reflected by a succession of NCCAs/CCAs cycles) [19–21]. This key point has been ignored by the current gene mutation based cancer

theory. In order to correctly understand the mechanism of cancer, the importance of the two phases of evolutionary dynamics must be appreciated, as the two phases of dynamics should be detectable at levels below (such as the gene level) and above the genome [21]. It would be interesting to investigate what patterns of karyotype evolution occur in other common diseases. Based on our studies on different stress conditions inducing elevated levels of genome alteration, it is reasonable to link genome alterations to many common diseases. Increasing numbers of reports are pointing in this direction [7,74,75]. However, there needs to be a greater emphasis on stochastic genome alterations rather than clonal somatic genome variations.

4.4. Integrating the time factor in genomic research

The evolutionary approach in genomic studies must integrate the time factor as time is a key aspect of the evolutionary process. Such studies require sampling the process from multiple time windows in order to watch evolution in action. As evolution is a stochastic process, and each tumor represents one run of independent evolution, the molecular pathways involved in any given tumor might be drastically different from one another. This fact might be the greatest challenge when trying to apply individual pathway information to patient care. However, by applying evolutionary principles, the evolutionary potential may be predictable if based on the overall system stability. For example, by ignoring individual specific molecular mechanisms, we have focused on measuring the frequencies of NCCAs, which is an indicator of system dynamics, to study the evolutionary potential of cancer. Regardless of what the contributing factors are, as soon as the frequencies of NCCAs reached a certain level, cancer formation happened in due time. Further research needs to integrate the degree of system instability and its progression over time. There are some studies that address this issue [76,77], however they have been focused on gene mutations rather than genome alterations.

5. Conclusion and future perspective

Due to the confusion between the genome context and the gene content of the genome, a fashionable trend of pushing genomic research "beyond the genome" has started, following the sequencing phase of the genome project. Beyond the genome has become a topic for genomic conferences and popular science. The truth of the matter is that we have not yet decoded the genome and do not understand how the genome defines the genetic network. Sequencing the whole genome is just the first step to decoding the genome's DNA, and the genome's crucial topological role and the mechanism of how the different parts interact and create emergent functional properties is basically unknown. It is important to realize that various networks represent the emergent properties of the whole genome rather than limited genes. Equally important, altered genomes lead to altered network structures. It is thus less useful to study network dynamics without monitoring the genome status. The real task for the research community is to begin to really study the genome rather than look beyond it. To achieve the goal of decoding the genome, new fundamental principles of biology must be formed to provide a conceptual framework and the needed guidance to develop new technological platforms. Only these new genome based concepts and methodologies (rather than the DNA technologies that we have used so far) can effectively study genome level dynamics and alterations as well as implications to human diseases.

This year marks the tenth anniversary of the completion of the sequencing of the human genome (Human Genome Project) [4,9,10]. Since the launching of this project in 1990, both basic knowledge and genomic technologies have drastically grown (e.g. from Sanger-based capillary sequencing to massive parallel sequencing; from monitoring the expression of a couple of genes to global expression analysis). There is now an increased appreciation of important features of the human genome, including the number of genes, AT/GC-content, the high degree of alternative splicing (many more proteins than genes), gene order/cluster, conservation of genes and non-coding sequences, the similarity of the proteome across placental mammals (where the creation of fundamental new proteins is rare), the significance of transposons and non-coding RNA, copy number variations, linkage disequilibrium, syntenic genomic blocks among species, codon usage bias, and many others. These new discoveries/technologies have also been applied to human diseases studies, where a total of 2850 genes contributing to Mendelian diseases have been identified, and as more than 1100 loci affecting more than 165 diseases have been associated with common diseases, this illustrates the high levels of genetic heterogeneity among patients [9]. Paradoxically, the success of identifying large numbers of genetic factors poses the greatest challenge for their application to clinical use. For instance, there are over 10,000 different genetic variants that have been associated with Schizophrenia. Each of them is relatively rare and responsible for a tiny increase in disease risk [78]. What is the clinical significance of such diverse genetic variants? Despite these impressive findings, little is known on how to decode the genome as we have discussed. For example, how does self-organization act upon the genome context? How can the genomic topology be revealed from sequencing data? How do altered genomes affect genetic networks and what is the relationship between genome reorganization and re-wiring? What is the relationship among different levels of genetic systems and the implication to human diseases? Are genome level alterations responsible for the missing heritability? Can massive amounts of genomic data be applied to clinical settings? Despite many positive comments regarding the impact of the sequencing project, a key conclusion that is emerging is that just sequencing DNA will not reveal the mystery of life, and a new way of thinking that departs from the gene theory is now urgently needed [7,13,15,27,31].

The first and hardest step is accepting the fact that decoding the genome and sequencing DNA are significantly different, and the genomic landscape will not be revealed simply by sequencing more DNA samples [79]. It is time for the research community to develop new concepts and technologies, rather than to continue to pile on more of what we already know. We have recently introduced the genome theory as a departure from the gene theory [21,24,27]. There has also been an increasing call in chromosomal studies that "a DNA sequence isn't enough: to understand the workings of the genome, we must study chromosome structure" [80].

To advance the transition from gene oriented research to genome based research, the following avenues need to be explored. First, technical platforms are required to study system behavior (e.g. evolutionary dynamics) without falling back on traditional approaches that simply characterize the low level parts. For example, to monitor the evolutionary potential of cancer evolution, it is more reliable to measure genome level heterogeneity rather than monitoring each specific individual genetic defect or pathway, as there are potentially so many defects that attempt to analyze them all results in fuzzy data sets with little or no meaning. By focusing on a different level, the lower level of seemingly infinite complexity can sometimes be converted into a higher level of simplicity [31].

Second, systematic comparisons of the relationships of the genome, transcriptome and proteome are needed [81,82]. Much as the complement of all genes is not equal to the genome, the proteins of a given genome are not the entire proteome without the essential components of interaction dynamics (where the cellular topology is important and is related to genome topology). Similarly, the viewpoint that the proteome is larger than the genome (based on there being more proteins than genes) is not accurate. Again, it is not only the number of protein species that matter, but the interaction potential that exists which is defined by the genome. Sequencing the genome only reveals the code of the parts list. The majority of current proteomics research analyzes only the parts and its potential interactions are artificially

constrained and defined by the controlled experiment. Without real cell topology, the averaged interaction map is artificial and not truly representative of an actual living system and therefore of limited clinical value. Thus, the transcriptome and proteome should be considered as emergent properties of the genome. The understanding of the convergent and divergent relationship among them is of importance, especially the similarity and difference between typical Mendelian diseases and common diseases, and between physiological, pathological and medical interventional conditions.

Third, there are many fascinating studies on network re-wiring when altering master regulators [83,84]. Further analyses are needed to integrate genome topology and compare network re-wiring within a given genome and to determine how the formation of a new genome re-creates a new boundary for the new network, as they involve different patterns of evolutionary dynamics. Clinically useful quantitative models are needed with prediction value. So far, some big challenges for system biology (where the gene theory dominates) include dealing with multiple levels of heterogeneity that occur in real biological systems (where the genome theory is essential) and applying them to design practical useful models.

Fourth, genome chaos represents a powerful model system to understand the gene and genome relationship, and their contributions to human diseases [18,31]. With various experimental manipulations, the evolutionary process of a chaotic genome can be traced using single cell genomic analysis to study both the micro and macro phases of evolution. In addition, networks can be systematically compared within physiological and pathological conditions, including with or without drug treatments. More importantly, it should finally be determined whether quantitative features from the gene level can be used to predict behavior occurring at the genome level, which could also offer insight on how information converges and/or diverges at different levels.

Fifth, cytogenomic analysis should once again be restored as a uniquely advantageous method to analyze the genome at the karyotype level and is capable of studying individual cells as well as cell populations. We are not recommending the current trend of the molecularization of current cytogenetic analysis, as copy number variation testing cannot be used to replace karyotype analysis (not only do karyoytypic and sub-chromosomal changes represent different levels of a system, but current molecular technologies profile the average, favoring the biased clonal populations while eliminating important heterogeneity). Similarly, to focus solely on individual chromatin domains is not sufficient to study genome topology, the self-organization principle or the patterns of system dynamics and their potential interactions. New chromosomal research needs to stem from a genome based view rather than a gene based view or limited chromatin domain view [85]. Clearly, these exciting technologies and observations from chromatin loop domains need to be integrated into the level of the whole genome package [43–45,86–88]. Similarly, chromosome based research should not only lead to explanations of a genetic link to specific genes but should also consider the genome perspective. For example, the three dimensional synteny among different species needs to be analyzed. Another important area to investigate is the linkage between karyotypes and species to determine whether species display different genome context and thereby establish the fact that it is the genome rather genes that defines a species. Such an analysis should also be carried out in cancer research to illustrate that the key feature of cancer is non-shared random genome alterations.

Expectations of future genomic research need to be adjusted as well, as there has been a genomic bubble in terms of its clinical promise [14]. Due to the high degree of complexity and heterogeneity of the multiple levels of the genetic system, one cannot always determine a distinct causative relationship. Our recent data illustrates that there might be no clear causative relationship at the molecular level within a complex biosystem (31, unpublished data). Common diseases might not have dominant common patterns similar to those identified in typical Mendelian disease [89]. It is likely that stochastic genome alterations are involved in common diseases and are the general rule rather than the exception [7]. Our studies have made the case in cancer research and now it is time to apply the genome theory to studies of other common diseases [5,7,24,27,75].

Another important realization is that communication with the non-scientific community about the clinical implications of genomic information needs to be more realistic. Yes, we can sequence tens of thousands of cancer samples and establish the most comprehensive genomic catalogs with highly heterogeneous details, but this information might not be useful clinically for prediction and treatment of individual patients, as “there are likely to be fundamental limits on precise prediction due to the complex architecture of common traits” [9]. If high heterogeneity of the genome is a key for many common diseases, sequencing DNA will not deliver the promise of major medical advancement. Another important note to large scale genomic research is that in addition to paying attention to statistical significance, there also is a huge need to determine the biological relevance.

A powerful approach for future functional genomics is to apply the evolutionary principle and methodologies to study the relationship between genome alteration, copy number variation, gene mutation, epigenetic variation and network dynamics. We predict that the identification of genome level alterations is the most important level of a system, this will lead to the understanding that the overall genome system and its alterations are more significant than the lower level parts of that system [27,30,32,90,91,92,93,94]. In particular, a multiple dimension global interactive map is needed exhibiting important transitions between germline cells to various types of somatic cells, from physiological conditions to pathological conditions, and demonstrating differences between normal and abnormal genome systems.

Decoding the genome is much more difficult than analyzing DNA as despite our superior capability to sequence DNA, we have not achieved a fundamental technical platform to understand genome level coding. However, this difficult transition MUST be achieved in order to push the field of genomics forward. Welcome to the challenges of the Genome Age.

Acknowledgments

We dedicate this perspective to Dr. Lap-Chee Tsui in honor of his 60th birthday. HH deeply appreciates Dr. Tsui's outstanding mentorship and influence. This manuscript is part of a series of studies entitled, “The mechanisms of somatic cell and organismal evolution”. We would like to thank Gloria Heppner; Gary Stein; Donald Coffey and O.J. Miller for their continuous support and interest in this project. Thanks also go to Markku Kurkinen, Leonard Lipovich, Rafe Furst, Steve Krawetz and Jian-Bing Fan for discussions. This work was partially supported by grants from the Susan G. Komen Breast Cancer Foundation, SeeDNA Inc and the DOD (GW093028).

References

[1] T.A. Pearson, T.A. Manolio, How to interpret a genome-wide association study, JAMA 299 (2008) 1335–1344.
[2] O. Alter, P.O. Brown, D. Botstein, Singular value decomposition for genome-wide expression data processing and modeling, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 10101–10106.
[3] L. Feuk, A.R. Carson, S.W. Scherer, Structural variation in the human genome, Nat. Rev. Genet. 7 (2006) 85–97.
[4] E.R. Mardis, A decade's perspective on DNA sequencing technology, Nature 470 (2011) 198–203.
[5] H.H. Heng, Cancer genome sequencing: the challenges ahead, Bioessays 29 (2007) 783–794.
[6] T.A. Manolio, F.S. Collins, N.J. Cox, D.B. Goldstein, L.A. Hindorff, D.J. Hunter, M.I. McCarthy, E.M. Ramos, L.R. Cardon, A. Chakravarti, J.H. Cho, A.E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C.N. Rotimi, M. Slatkin, D. Valle, A.S. Whittemore, M. Boehnke, A.G. Clark, E.E. Eichler, G. Gibson, J.L. Haines, T.F. Mackay, S.A. McCarroll,

P.M. Visscher, Finding the missing heritability of complex diseases, Nature 461 (2009) 747–753.
[7] H.H. Heng, Missing heritability and stochastic genome alterations, Nat. Rev. Genet. 11 (2010) 813.
[8] H.H. Heng, S.W. Bremer, J.B. Stevens, K.J. Ye, G. Liu, C.J. Ye, Genetic and epigenetic heterogeneity in cancer: a genome-centric perspective, J. Cell. Physiol. 220 (2009) 538–547.
[9] E.S. Lander, Initial impact of the sequencing of the human genome, Nature 470 (2011) 187–197.
[10] E.D. Green, M.S. Guyer, Charting a course for genomic medicine from base pairs to bedside, Nature 470 (2011) 204–213.
[11] L.D. Wood, D.W. Parsons, S. Jones, J. Lin, T. Sjoblom, R.J. Leary, D. Shen, S.M. Boca, T. Barber, J. Ptak, N. Silliman, S. Szabo, Z. Dezso, V. Ustyanksky, T. Nikolskaya, Y. Nikolsky, R. Karchin, P.A. Wilson, J.S. Kaminker, Z. Zhang, R. Croshaw, J. Willis, D. Dawson, M. Shipitsin, J.K. Willson, S. Sukumar, K. Polyak, B.H. Park, C.L. Pethiyagoda, P.V. Pant, D.G. Ballinger, A.B. Sparks, J. Hartigan, D.R. Smith, E. Suh, N. Papadopoulos, P. Buckhaults, S.D. Markowitz, G. Parmigiani, K.W. Kinzler, V.E. Velculescu, B. Vogelstein, The genomic landscapes of human breast and colorectal cancers, Science 318 (2007) 1108–1113.
[12] M.F. Berger, M.S. Lawrence, F. Demichelis, Y. Drier, K. Cibulskis, A.Y. Sivachenko, A. Sboner, R. Esgueva, D. Pflueger, C. Sougnez, R. Onofrio, S.L. Carter, K. Park, L. Habegger, L. Ambrogio, T. Fennell, M. Parkin, G. Saksena, D. Voet, A.H. Ramos, T.J. Pugh, J. Wilkinson, S. Fisher, W. Winckler, S. Mahan, K. Ardlie, J. Baldwin, J.W. Simons, N. Kitabayashi, T.Y. MacDonald, P.W. Kantoff, L. Chin, S.B. Gabriel, M.B. Gerstein, T.R. Golub, M. Meyerson, A. Tewari, E.S. Lander, G. Getz, M.A. Rubin, L.A. Garraway, The genomic complexity of primary human prostate cancer, Nature 470 (2011) 214–220.
[13] E.E. Eichler, J. Flint, G. Gibson, A. Kong, S.M. Leal, J.H. Moore, J.H. Nadeau, Missing heritability and strategies for finding the underlying causes of complex disease, Nat. Rev. Genet. 11 (2010) 446–450.
[14] J.P. Evans, E.M. Meslin, T.M. Marteau, T. Caulfield, Genomics. Deflating the genomic bubble, Science 331 (2011) 861–862.
[15] H.H. Heng, The conflict between complex systems and reductionism, JAMA 300 (2008) 1580–1581.
[16] H.H. Heng, G. Liu, J.B. Stevens, S.W. Bremer, K.J. Ye, C.J. Ye, Genetic and epigenetic heterogeneity in cancer: the ultimate challenge for drug therapy, Curr. Drug Targets 11 (2010) 1304–1316.
[17] H.H. Heng, J.B. Stevens, G. Liu, S.W. Bremer, C.J. Ye, Imaging genome abnormalities in cancer research, Cell Chromosome 3 (2004) 1.
[18] H.H. Heng, J.B. Stevens, G. Liu, S.W. Bremer, K.J. Ye, P.V. Reddy, G.S. Wu, Y.A. Wang, M.A. Tainsky, C.J. Ye, Stochastic cancer progression driven by non-clonal chromosome aberrations, J. Cell. Physiol. 208 (2006) 461–472.
[19] H.H. Heng, G. Liu, S. Bremer, K.J. Ye, J. Stevens, C.J. Ye, Clonal and non-clonal chromosome aberrations and genome variation and aberration, Genome 49 (2006) 195–204.
[20] H.H. Heng, S.W. Bremer, J. Stevens, K.J. Ye, F. Miller, G. Liu, C.J. Ye, Cancer progression by non-clonal chromosome aberrations, J. Cell. Biochem. 98 (2006) 1424–1435.
[21] C.J. Ye, G. Liu, S.W. Bremer, H.H. Heng, The dynamics of cancer chromosomes and genomes, Cytogenet. Genome Res. 118 (2007) 237–246.
[22] H.H. Heng, J.B. Stevens, L. Lawrenson, G. Liu, K.J. Ye, S.W. Bremer, C.J. Ye, Patterns of genome dynamics and cancer evolution, Cell. Oncol. 30 (2008) 513–514.
[23] H.H. Heng, The gene-centric concept: a new liability? Bioessays 30 (2008) 196–197.
[24] H.H. Heng, J.B. Stevens, S.W. Bremer, K.J. Ye, G. Liu, C.J. Ye, The evolutionary mechanism of cancer, J. Cell. Biochem. 109 (2010) 1072–1084.
[25] P. Duesberg, Chromosomal chaos and cancer, Sci. Am. 296 (2007) 52–59.
[26] J.M. Nicholson, P. Duesberg, On the karyotypic origin and evolution of cancer cells, Cancer Genet. Cytogenet. 194 (2009) 96–110.
[27] H.H. Heng, The genome-centric concept: resynthesis of evolutionary theory, Bioessays 31 (2009) 512–525.
[28] http://www.genome.gov/10005107.
[29] American Association for Cancer Research Human Epigenome Task Force, European Union, Network of Excellence, Scientific Advisory Board, Moving AHEAD with an international human epigenome project, Nature 454 (2008) 711–715.
[30] P.M. Durand, R.E. Michod, Genomics in the light of evolutionary transitions, Evolution 64 (2010) 1533–1540.
[31] H.H.Q. Heng, Bio-complexity: challenging reductionism, Handbook on Systems and Complexity in Health, (in press).
[32] B. McClintock, The significance of responses of the genome to challenge, Science 226 (1984) 792–801.
[33] F. Crick, What mad pursuit: a personal view of scientific discovery, Basic Books, New York, 1988.
[34] M. Srivastava, O. Simakov, J. Chapman, B. Fahey, M.E. Gauthier, T. Mitros, G.S. Richards, C. Conaco, M. Dacre, U. Hellsten, C. Larroux, N.H. Putnam, M. Stanke, M. Adamska, A. Darling, S.M. Degnan, T.H. Oakley, D.C. Plachetzki, Y. Zhai, M. Adamski, A. Calcino, S.F. Cummins, D.M. Goodstein, C. Harris, D.J. Jackson, S.P. Leys, S. Shu, B.J. Woodcroft, M. Vervoort, K.S. Kosik, G. Manning, B.M. Degnan, D.S. Rokhsar, The Amphimedon queenslandica genome and the evolution of animal complexity, Nature 466 (2010) 720–726.
[35] J.K. Colbourne, M.E. Pfrender, D. Gilbert, W.K. Thomas, A. Tucker, T.H. Oakley, S. Tokishita, A. Aerts, G.J. Arnold, M.K. Basu, D.J. Bauer, C.E. Caceres, L. Carmel, C. Casola, J.H. Choi, J.C. Detter, Q. Dong, S. Dusheyko, B.D. Eads, T. Frohlich, K.A. Geiler-Samerotte, D. Gerlach, P. Hatcher, S. Jogdeo, J. Krijgsveld, E.V. Kriventseva, D. Kultz, C. Laforsch, E. Lindquist, J. Lopez, J.R. Manak, J. Muller, J. Pangilinan, R.P. Patwardhan, S. Pitluck, E.J. Pritham, A. Rechtsteiner, M. Rho, I.B. Rogozin, O. Sakarya, A. Salamov, S. Schaack, H. Shapiro, Y. Shiga, C. Skalitzky, Z. Smith, A. Souvorov, W. Sung, Z. Tang, D. Tsuchiya, H. Tu, H. Vos, M. Wang, Y.I. Wolf, H. Yamagata, T. Yamada, Y. Ye, J.R. Shaw, J. Andrews, T.J. Crease, H. Tang, S.M. Lucas, H.M. Robertson, P. Bork, E.V. Koonin, E.M. Zdobnov, I.V. Grigoriev, M. Lynch, J.L. Boore, The ecoresponsive genome of Daphnia pulex, Science 331 (2011) 555–561.
[36] M. Kohn, J. Hogel, W. Vogel, P. Minich, H. Kehrer-Sawatzki, J.A. Graves, H. Hameister, Reconstruction of a 450-My-old ancestral vertebrate protokaryotype, Trends Genet. 22 (2006) 203–210.
[37] M.A. Ferguson-Smith, V. Trifonov, Mammalian karyotype evolution, Nat. Rev. Genet. 8 (2007) 950–962.
[38] W.J. Murphy, D.M. Larkin, A. Everts-van der Wind, G. Bourque, G. Tesler, L. Auvil, J.E. Beever, B.P. Chowdhary, F. Galibert, L. Gatzke, C. Hitte, S.N. Meyers, D. Milan, E.A. Ostrander, G. Pape, H.G. Parker, T. Raudsepp, M.B. Rogatcheva, L.B. Schook, L.C. Skow, M. Welge, J.E. Womack, J. O'Brien S, P.A. Pevzner, H.A. Lewin, Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps, Science 309 (2005) 613–617.
[39] H.H. Heng, Elimination of altered karyotypes by sexual reproduction preserves species identity, Genome 50 (2007) 517–524.
[40] A.S. Wilkins, R. Holliday, The evolution of meiosis from mitosis, Genetics 181 (2009) 3–12.
[41] H.H. Heng, J.W. Chamberlain, X.M. Shi, B. Spyropoulos, L.C. Tsui, P.B. Moens, Regulation of meiotic chromatin loop size by chromosomal position, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 2795–2800.
[42] J.M. Palmer, N.P. Keller, Secondary metabolism in fungi: does chromosomal location matter? Curr. Opin. Microbiol. 13 (2010) 431–436.
[43] H.H. Heng, S.A. Krawetz, W. Lu, S. Bremer, G. Liu, C.J. Ye, Re-defining the chromatin loop domain, Cytogenet. Cell Genet. 93 (2001) 155–161.
[44] H.H. Heng, S. Goetze, C.J. Ye, G. Liu, J.B. Stevens, S.W. Bremer, S.M. Wykes, J. Bode, S.A. Krawetz, Chromatin loops are selectively anchored using scaffold/matrix-attachment regions, J. Cell Sci. 117 (2004) 999–1008.
[45] J. Bode, S. Goetze, H. Heng, S.A. Krawetz, C. Benham, From DNA structure to gene expression: mediators of nuclear compartmentalization and dynamics, Chromosome Res. 11 (2003) 435–445.
[46] R. Gorelick, H.H. Heng, Sex reduces genetic variation: a multidisciplinary review, Evolution 65 (2011) 1088–1098.
[47] D.J. Futuyma, Evolutionary constraint and ecological consequences, Evolution 64 (2010) 1865–1884.
[48] K. Kitada, A. Taima, K. Ogasawara, S. Metsugi, S. Aikawa, Chromosome-specific segmentation revealed by structural analysis of individually isolated chromosomes, Genes Chromosomes Cancer 50 (2011) 217–227.
[49] J.R. Lupski, Genomic rearrangements and sporadic disease, Nat. Genet. 39 (2007) S43–S47.
[50] I.Y. Iourov, S.G. Vorsanova, Y.B. Yurov, Chromosomal mosaicism goes global, Mol Cytogenet 1 (2008) 26.
[51] J.H. Bielas, K.R. Loeb, B.P. Rubin, L.D. True, L.A. Loeb, Human cancers express a mutator phenotype, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 18238–18242.
[52] D.H. Erwin, E.H. Davidson, The evolution of hierarchical gene regulatory networks, Nat. Rev. Genet. 10 (2009) 141–148.
[53] S. Huang, I. Ernberg, S. Kauffman, Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective, Semin. Cell Dev. Biol. 20 (2009) 869–876.
[54] P. Ao, Global view of bionetwork dynamics: adaptive landscape, J. Genet. Genomics 36 (2009) 63–73.
[55] H.H. Heng, B. Spyropoulos, P.B. Moens, FISH technology in chromosome and genome research, Bioessays 19 (1997) 75–84.
[56] C. Lanctot, T. Cheutin, M. Cremer, G. Cavalli, T. Cremer, Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions, Nat. Rev. Genet. 8 (2007) 104–115.
[57] J.J. Roix, P.G. McQueen, P.J. Munson, L.A. Parada, T. Misteli, Spatial proximity of translocation-prone gene loci in human lymphomas, Nat. Genet. 34 (2003) 287–291.
[58] T. Misteli, The inner life of the genome, Sci. Am. 304 (2011) 66–73.
[59] E. Lieberman-Aiden, N.L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B.R. Lajoie, P.J. Sabo, M.O. Dorschner, R. Sandstrom, B. Bernstein, M.A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L.A. Mirny, E.S. Lander, J. Dekker, Comprehensive mapping of long-range interactions reveals folding principles of the human genome, Science 326 (2009) 289–293.
[60] N.L. van Berkum, E. Lieberman-Aiden, L. Williams, M. Imakaev, A. Gnirke, L.A. Mirny, J. Dekker, E.S. Lander, Hi-C: a method to study the three-dimensional architecture of genomes, J. Vis. Exp. (2010).
[61] D. Bau, A. Sanyal, B.R. Lajoie, E. Capriotti, M. Byron, J.B. Lawrence, J. Dekker, M.A. Marti-Renom, The three-dimensional folding of the alpha-globin gene domain reveals formation of chromatin globules, Nat. Struct. Mol. Biol. 18 (2011) 107–114.
[62] A.M. Hillmer, F. Yao, K. Inaki, W.H. Lee, P.N. Ariyaratne, A.S. Teo, X.Y. Woo, Z. Zhang, H. Zhao, L. Ukil, J.P. Chen, F. Zhu, J.B. So, M. Salto-Tellez, W.T. Poh, K.F. Zawack, N. Nagarajan, S. Gao, G. Li, V. Kumar, H.P. Lim, Y.Y. Sia, C.S. Chan, S.T. Leong, S.C. Neo, P.S. Choi, H. Thoreau, P.B. Tan, A. Shaha, X. Ruan, J. Bergh, P. Hall, V. Cacheux-Rataboul, C.L. Wei, K.G. Yeoh, W.K. Sung, G. Bourque, E.T. Liu, Y. Ruan, Comprehensive long-span paired-end-tag mapping reveals characteristic patterns of structural variations in epithelial cancer genomes, Genome Res. 21 (2011) 665–675.
[63] H.H. Heng, C.J. Ye, F. Yang, S. Ebrahim, G. Liu, S.W. Bremer, C.M. Thomas, J. Ye, T.J. Chen, C. Tuck-Muller, J.W. Yu, S.A. Krawetz, A. Johnson, Analysis of marker or complex chromosomal rearrangements present in pre- and post-natal karyotypes

utilizing a combination of G-banding, spectral karyotyping and fluorescence in situ hybridization, Clin. Genet. 63 (2003) 358–367.
[64] C.J. Ye, J.B. Stevens, G. Liu, S.W. Bremer, A.S. Jaiswal, K.J. Ye, M.F. Lin, L. Lawrenson, W.D. Lancaster, M. Kurkinen, J.D. Liao, C.G. Gairola, M.P. Shekhar, S. Narayan, F.R. Miller, H.H. Heng, Genome based cell population heterogeneity promotes tumorigenicity: the evolutionary mechanism of cancer, J. Cell. Physiol. 219 (2009) 288–300.
[65] E. Schrock, S. du Manoir, T. Veldman, B. Schoell, J. Wienberg, M.A. Ferguson-Smith, Y. Ning, D.H. Ledbetter, I. Bar-Am, D. Soenksen, Y. Garini, T. Ried, Multicolor spectral karyotyping of human chromosomes, Science 273 (1996) 494–497.
[66] C.J. Ye, W. Lu, G. Liu, S.W. Bremer, Y.A. Wang, P. Moens, M. Hughes, S.A. Krawetz, H.H. Heng, The combination of SKY and specific loci detection with FISH or immunostaining, Cytogenet. Cell Genet. 93 (2001) 195–202.
[67] J.B. Stevens, G. Liu, S.W. Bremer, K.J. Ye, W. Xu, J. Xu, Y. Sun, G.S. Wu, S. Savasan, S.A. Krawetz, C.J. Ye, H.H. Heng, Mitotic cell death by chromosome fragmentation, Cancer Res. 67 (2007) 7686–7694.
[68] J.B. Stevens, B.Y. Abdallah, S.M. Regan, G. Liu, S.W. Bremer, C.J. Ye, H.H. Heng, Comparison of mitotic cell death by chromosome fragmentation to premature chromosome condensation, Mol Cytogenet 3 (2010) 20.
[69] P.J. Stephens, C.D. Greenman, B. Fu, F. Yang, G.R. Bignell, L.J. Mudie, E.D. Pleasance, K.W. Lau, D. Beare, L.A. Stebbings, S. McLaren, M.L. Lin, D.J. McBride, I. Varela, S. Nik-Zainal, C. Leroy, M. Jia, A. Menzies, A.P. Butler, J.W. Teague, M.A. Quail, J. Burton, H. Swerdlow, N.P. Carter, L.A. Morsberger, C. Iacobuzio-Donahue, G.A. Follows, A.R. Green, A.M. Flanagan, M.R. Stratton, P.A. Futreal, P.J. Campbell, Massive genomic rearrangement acquired in a single catastrophic event during cancer development, Cell 144 (2011) 27–40.
[70] M. Meyerson, D. Pellman, Cancer genomes evolve by pulverizing single chromosomes, Cell 144 (2011) 9–10.
[71] H.Q. Heng, W.Y. Chen, Y.C. Wang, Effects of pingyanymycin on chromosomes: a possible structural basis for chromosome aberration, Mutat. Res. 199 (1988) 199–205.
[72] H.H. Heng, J. Squire, L.C. Tsui, High-resolution mapping of mammalian genes by in situ hybridization to free chromatin, Proc. Natl. Acad. Sci. U.S.A. 89 (1992) 9509–9513.
[73] G. Rancati, N. Pavelka, B. Fleharty, A. Noll, R. Trimble, K. Walton, A. Perera, K. Staehling-Hampton, C.W. Seidel, R. Li, Aneuploidy underlies rapid adaptive evolution of yeast cells deprived of a conserved cytokinesis motor, Cell 135 (2008) 879–893.
[74] C.L. Smith, A. Bolton, G. Nguyen, Genomic and epigenomic instability, fragile sites, schizophrenia and autism, Curr Genomics 11 (2010) 447–469.
[75] I.Y. Iourov, S.G. Vorsanova, Y.B. Yurov, Somatic genome variations in health and disease, Curr Genomics 11 (2010) 387–396.
[76] C.C. Maley, P.C. Galipeau, J.C. Finley, V.J. Wongsurawat, X. Li, C.A. Sanchez, T.G. Paulson, P.L. Blount, R.A. Risques, P.S. Rabinovitch, B.J. Reid, Genetic clonal diversity predicts progression to esophageal adenocarcinoma, Nat. Genet. 38 (2006) 468–473.
[77] S. Jones, W.D. Chen, G. Parmigiani, F. Diehl, N. Beerenwinkel, T. Antal, A. Traulsen, M.A. Nowak, C. Siegel, V.E. Velculescu, K.W. Kinzler, B. Vogelstein, J. Willis, S.D. Markowitz, Comparative lesion sequencing provides insights into tumor evolution, Proc. Natl. Acad. Sci. U.S.A. 105 (2008) 4283–4288.
[78] N. Wade, Hoopla, and disappointment, in schizophrenia research, The New York Times, The New York Times Company, New York, 2009.
[79] H.H. Heng, Joshua B. Stevens, Steve W. Bremer, Guo Liu, Batoul Y. Abdallah, Christine J. Ye, Evolutionary mechanisms and diversity in cancer, Adv Cancer Res (in press).
[80] M. Baker, Genomics: genomes in three dimensions, Nature 470 (2011) 289–294.
[81] L. Lipovich, R. Johnson, C.Y. Lin, MacroRNA underdogs in a microRNA world: evolutionary, regulatory, and biomedical significance of mammalian long non-protein-coding RNA, Biochim. Biophys. Acta 1799 (2010) 597–615.
[82] P. Carninci, Y. Hayashizaki, Noncoding RNA transcription beyond annotated genes, Curr. Opin. Genet. Dev. 17 (2007) 139–144.
[83] M. Isalan, C. Lemerle, K. Michalodimitrakis, C. Horn, P. Beltrao, E. Raineri, M. Garriga-Canut, L. Serrano, Evolvability and hierarchy in rewired bacterial gene networks, Nature 452 (2008) 840–845.
[84] S.R. Paladugu, S. Zhao, A. Ray, A. Raval, Mining protein networks for synthetic genetic interactions, BMC Bioinforma. 9 (2008) 426.
[85] D. Carter, L. Chakalova, C.S. Osborne, Y.F. Dai, P. Fraser, Long-range chromatin regulatory interactions in vivo, Nat. Genet. 32 (2002) 623–626.
[86] D. Ottaviani, E. Lever, R. Mitter, T. Jones, T. Forshew, R. Christova, E.M. Tomazou, V.K. Rakyan, S.A. Krawetz, A.E. Platts, B. Segarane, S. Beck, D. Sheer, Reconfiguration of genomic anchors upon transcriptional activation of the human major histocompatibility complex, Genome Res. 18 (2008) 1778–1786.
[87] E. Chevret, E.V. Volpi, D. Sheer, Mini review: form and function in the human interphase chromosome, Cytogenet. Cell Genet. 90 (2000) 13–21.
[88] S. Horike, S. Cai, M. Miyano, J.F. Cheng, T. Kohwi-Shigematsu, Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome, Nat. Genet. 37 (2005) 31–40.
[89] J. McClellan, M.C. King, Genetic heterogeneity in human diseases, Cell 141 (2010) 210–217.
[90] R. Gorelick, R.M.D. Laubichler, Genetic = Heritable (Genetic ≠ DNA), Biol Theory 3 (2008) 79–84.
[91] J.B. Stevens, B.Y. Abdallah, G. Liu, C.J. Ye, S.D. Horne, G. Wang, S. Savasan, M. Shekhar, S.A. Krawetz, M. Hüttemann, M.A. Tainsky, G.S. Wu, Y. Xie, K. Zhang, H.Q. Heng, Diverse system stresses: common mechanisms of chromosome fragmentation, Cell Death Disease (in press).
[92] P.A. Astolfi, F. Salamini, V. Sgaramella, Are we Genomic Mosaics? Variations of the Genome of Somatic Cells can Contribute to Diversify our Phenotypes, Curr. Genomics 11 (2010) 379–386.
[93] M.S. Abu-Asab, M. Chaouchi, S. Alesci, S. Galli, M. Laassri, A.K. Cheema, F. Atouf, J. VanMeter, H. Amri, Biomarkers in the age of omics: time for a systems biology approach, Omics 15 (2011) 105–112.
[94] D.R. Forsdyke, Scherrer and Jost's symposium: the gene concept in 2008, Theory Biosci. 128 (2009) 157–161.